Big Data Research 3 (2016) 2–9

Contents lists available at ScienceDirect

Big Data Research

www.elsevier.com/locate/bdr

SFPM: A Secure and Fine-Grained Privacy-Preserving Matching Protocol for Mobile Social Networking ☆,☆☆

Xue Yang [a,*], Rongxing Lu [b], Hongbin Liang [c], Xiaohu Tang [a]

[a] The Information Security and National Computing Grid Laboratory, Southwest Jiaotong University, Chengdu, 610031, China
[b] School of Electrical and Electronic Engineering, Nanyang Technological University, 50 Nanyang Avenue, 639798, Singapore
[c] School of Transportation and Logistics, Southwest Jiaotong University, Chengdu, 610031, China

ARTICLE INFO

Article history:
Received 30 May 2015
Received in revised form 9 October 2015
Accepted 3 November 2015
Available online 11 November 2015

Keywords:
Mobile social network
Big data
Proximity-based
Profile matching
Privacy preservation
Fine-grained

ABSTRACT

In emerging big data era, mobile social networking (MSN) is an important data source, which provides an attractive proximity-based communication platform for mobile users with similar interests, attributes, or background to communicate with each other. In this kind of proximity-based MSN, profile matching protocol, which enables a mobile user to break the ice and start a conversation with someone attractive, is one of important components for its success. However, profile matching may occasionally leak the sensitive information, hence privacy concerns often hinder users from enabling this functionality. Aiming at this problem, in this paper, we present a new secure and fine-grained privacy-preserving matching protocol, called SFPM. Differently from those previously reported private profile matching schemes, our proposed SFPM can fine-grainedly differentiate users with the same value of matching metrics by two phases of profile matching. In addition to the personal privacy preservation through secure and efficient cryptographic algorithm, SFPM also achieves the flexibility of profiles changing at the same time. Extensive performance evaluations via smartphones with android system are conducted, and experimental results demonstrate the effectiveness of the SFPM protocol.

1. Introduction

As mentioned by IBM, the rapid development of mobile social networking (MSN) shown in Fig. 1, promotes the generation of big data [1]. Actually, plentiful statistics have indicated that most of big data are produced by MSN, for example, the internet access records of Unicom users have reached 10 TB each day in China. Because of this rising situation, many applications based on big data mining and sharing, like the friend recommender systems of WeChat [2] and Twitter [3], and other personalized recommender systems [4–7], have been emerged. In these applications, when sharing the personal information, like location and preferences in public, people can receive a variety of useful location-based services from these recommender systems. In this paper, we focus on studying a kind of very popular location-based applications, called proximity-based friend recommendation (PFR) mentioned in [8], which allows physically proximate mobile users to have more tangible face-to-face social interaction in public places such as airports, trains and stadiums [9]. In general, one possible way is to use the widely known *profile matching* [10] technique, which is the first step to find the targeting user. As stated by Wu et al. [11], the essence of profile matching is that two users need to compare their personal profile attributes before real interaction. However, a real-world concern is that social profile attributes used in the profile matching process include sensitive information about users and the violation of the privacy of the users' social profiles may pose serious problems. Existing researches show that loss of privacy can expose users to unwanted advertisements [12] and spams/scams, cause social reputation or economic damage [13], and make them victims of blackmail or even physical violence [14]. Hence, the privacy concerns must be addressed when developing profile matching techniques for mobile social networks. In addition to security, clients of mobile social networks run on computing resource-constrained mobile devices. Therefore, a privacy-preserving and power-efficient profile matching scheme is needed for mobile social services.

Recently, there are quite a few schemes for private profile matching, which allow two users to compare their personal profiles without revealing private information to each other [10,15] have been researched. As mentioned in [16], there are two main-

☆ This article belongs to Big Data Networking.
☆☆ Fully documented templates are available in the elsarticle package on CTAN.
* Corresponding author.
E-mail addresses: xueyang.swjtu@gmail.com (X. Yang), rxlu@ntu.edu.sg (R. Lu), hbliang@home.swjtu.edu.cn (H. Liang), xhutang@swjtu.edu.cn (X. Tang).

http://dx.doi.org/10.1016/j.bdr.2015.11.001
2214-5796/

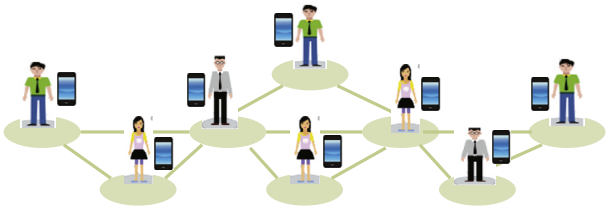

Fig. 1. Popular mobile social networking in big data era.

streams of approaches to solve the privacy-preserving profile-based friend matching problem. The first category treats the personal profile as a set of attributes and provides well-designed protocols based on private set intersection (PSI) and private cardinality of set intersection (PCSI) [10,17,18]; The second category considers the personal profile as a vector and measures the social proximity by private vector dot product or vector distance [19–22]. However, the vast majority of approaches in the first category have been proposed to enable only coarse-grained private matching and are unable to further differentiate users with the same attribute(s), which is less practical in applications [23]. To solve this problem and thus further enhance the usability of PFR in MSN, fine-grained private matching have been widely used in the second category, which are the basic idea of research in this paper. Hence, in what follows, we mainly discuss some related works of the second category.

Liang et al. proposed the multiple pseudonyms technique to improve the anonymity protection for profile matching protocol in [19], where secure dot-product computation is one of important building block. From the perspective of flexibility, multiple pseudonyms technique can ensure anonymity, but, it cannot satisfy the flexibility with slightly larger number of pseudonyms, which actually requires a lot of storage space and management overhead. In [21], Zhang et al. designed a fine-grained private matching protocol with different privacy levels in proximity-based mobile social networks, which included different matching metrics: l_1 distance and max distance. However, it did not consider the difference of profile items and is unable to further differentiate users with the same value of l_1 distance or the max distance. He et al. [24] addressed this issue by proposing a novel user self-controllable profile matching protocol, which allowed users to self-define the weighted of profile items during matching, thus provided more accurate matching results for users. Unfortunately, the method of matching information similarity in both [21] and [24] was based on the time-consuming paillier encryption [25] satisfying homogeneity. Thus, due to the heavy overheads of encryption and decryption, it is difficult to improve the overall operating of MSN applications. The purpose of this paper is to preserve private profile items from disclosing while improving the efficiency of schemes of the second category. In order to improve efficiency, we utilize some efficient methods to securely compute the vector dot product, while existing efficient methods are mainly two kinds. One is a new asymmetric scalar-product-preserving encryption proposed by Wong et al. [22], which is focused on the problem of k-nearest neighbor (kNN) computation on an encrypted database, however, it cannot satisfy the flexibility with the variation of profile items. The other is an efficient privacy-preserving cosine similarity computing (PPCSC) protocol proposed by Lu et al. [26], which could serve as the foundation of many research fields, like privacy-preserving big data miming, data access control, recommendation system. Extensive simulation results showed that the PPCSC protocol is the most efficient one in terms of computation and communication overheads. Thus, we choose the PPCSC protocol as the basis of our protocol. Moreover, most of privacy preserving profile matching protocols do not consider the attack model. To the best of our knowledge, none of the existing solutions to profile matching possesses all the desired properties: privacy-preserving, security (e.g., authentication and integrity), efficiency (e.g., cost-effective computation and communication overhead) and flexibility.

Therefore, how to achieve an efficient, flexible and privacy-preserving profile matching protocol is still challenging in proximity-based MSN. Aiming at the above challenge, in this paper, we propose a secure and finer-grained privacy-preserving matching protocol, called SFPM, for proximity-based MSN. With the SFPM protocol, users can efficiently and flexibly seek out the finer-grained matching target while without disclosing any personal information. In addition, our proposed protocol achieves the integrity of the message and source data authentication, and immensely decreases the computation overhead in comparison with that proposed in [21] and [24], especially alleviating the computational and communication burden of smartphones. Specifically, the main contributions of this paper are four aspects.

- We present SFPM, a new secure and fine-grained privacy-preserving matching protocol, which consists of two stages matching: cosine similarity and weighted l_1 norm. With SFPM, users can finer-grainedly distinguish users and find out the most matched one.
- Compared to the previous private matching protocols, SFPM provides a flexible and efficient matching style. In particular, we introduce a data processing center (DPC) to accomplish matching computations, which can immensely relieve the computation and communication burden of mobile devices. Moreover, the encryption algorithm proposed in [26] is more efficient and flexible compared with [22]. Consider the case when user inserts some profiles, only the inserted profiles should be encrypted, and then DPC only executes multiplication on these profiles and adds them in the previous computation result. Deleting and updating operations are similar with inserting. Therefore, our protocol is flexible for the variation of personal profiles.
- In addition to data confidentiality, the SFPM protocol achieves the integrity of the message and source data authentication by appending the message authentication codes, like the keyed-hashing for message authentication code (HMAC), as a result the ciphertexts can defense the additive noise.
- To validate the effectiveness of the proposed SFPM protocol, we implement both the SFPM protocol and the protocol one proposed by Zhang [21] on a platform with two android phones and a computer. By contrasting, we demonstrate that SFPM is much more efficient than existing similar profile matching schemes [21,24] in terms of the computational overhead.

The remainder of this paper is organized as follows. In Section 2, we formalize the system model and confirm the design goal. After that, we propose the SFPM protocol in Section 3. The security analysis and performance evaluation are introduced in Section 4 and 5, respectively. Finally, we draw our conclusions in Section 6.

2. System model and design goal

2.1. System model

In our system model, we consider a trusted key distribution center (KDC), a semi-trusted data precessing center (DPC), and a group of $l+1$ users $\mathbb{U} = \{U_A, U_1, U_2, \ldots, U_l\}$, where U_A is the requester of the PFR service and the others are the neighbors, as shown in Fig. 2, where we briefly represent the users with smartphones. KDC is a trustable and powerful entity, who is mainly